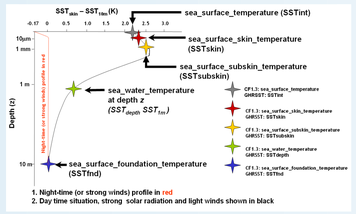
CLIMATE AND FORECAST (CF) CONVENTIONS FOR NETCDF -- THE FOUNDATION OF GODAE DATA INTEROPERABILITY

S. Hankin[1], J.D. Blower[2], T. Carval[3], K. Casey[4], C. Donlon[5], O. Lauret[6], T. Loubrieu[3], L. Petit de la Villeon[3], J. Piollé[3], S. Smith[7], A. Srinivasan[7], J. Trinanes[8], Ø. Godøy[9]

[1]*NOAA/Pacific Marine Environmental Laboratory, Seattle, US,* [2]*University of Reading Environmental Systems Science Centre, Reading, UK,* [3]*Ifremer , Brest, France,* [4]*NOAA/National Ocean Data Center,Washington D.C., US,* [5]*Met Office Hadley Centre International Group for High Resolution SST Project Office, Exeter, UK,* [6] *CLS Space Oceanography Division, Cedex, France,* [7]*Center for Ocean-Atmospheric Prediction Studies, The Florida State University, Talahassee, US,* [8]*NOAA/AOML (Miami,US) and University of Santiago de Compostela, Santiago de Compostela, Spain,* [9]*Norwegian Meteorological Institute*

GODAE High Resolution Sea Surface Temperature (SST) Pilot Project (GHRSST-PP)

The GODAE High Resolution Sea Surface Temperature (SST) Pilot Project (GHRSST-PP) was tasked with the generation of new satellite derived SST maps by merging complementary infrared and passive microwave satellite data. GHRSST-PP chose to use the CF convention because CF provided a well established community with a consensus orientation essential to the international partners of GHRSST-PP. CF was also in widespread use and provided extensive compatibility with interoperability servers. The choice of NetCDF/ CF for GHRSST-PP has resulted in a robust foundation for GHRSST-PP.



The hypothetical vertical profiles of temperature in low wind speed conditions during the night and day shown in the figure encapsulate the effects of the dominant heat transport processes and time scales of variability associated with distinct vertical and volume regimes (horizontal and temporal variability is implicitly assumed). CF-1.3 standard names and GHRSST short names are shown.

GHRSST-PP has actively promoted the use of NertCDF/CF as a baseline standard for products in the SST community Within CF, GHRSST-PP pioneered use of native satellite swath data as based on the application of conventional coordinate [lat,lon] attribute data (i.e. by treating swath data as irregular grids) and the use of binary coded flags (bitfields) using variable attributes. GHRSST-PP has developed and registered new SST definitions within the CF-1.3 standard_name table as shown in the Figure to the left.

Adherence to internationally agreed formats and interfaces gives agencies the confidence to invest in the development of GHRSST-PP data access and manipulation tools. This approach benefits both data users and producers, by simplifying data interfaces, international integration and federation of data, documentation, interoperability and reducing duplication of effort. Other groups can easily access and index GHRSST data and use of NetCDF/CF has resulted in a much better user service within GHRSST; as new satellite derived

What are the CF Conventions?

- A standard for data and metadata in netCDF files
- Encodes both model output and observational datasets

Examples of CF metadata
- Coordinate information needed to locate data in space and time
- Units, titles, standardized names for parameters
- Analytical grid information such as cell averaging methods

cfconventions.org

SAMOS
Shipboard Automated Meteorological and Oceanographic System Initiative

The SAMOS initiative (http://samos.coaps.fsu.edu/) collects, quality-controls, and distributes underway meteorological and near-surface oceanographic data from research vessels:
- Ship navigation: GPS position, course, heading, speed
- Air temperature, humidity, pressure, wind, solar radiation, etc.
- Sea temperature, salinity, conductivity, florescence

RV Knorr (has provided underway SAMOS data since 2005.

Cruises from 1 July 2005 - 30 Sept. 2008. Data are attached to emails transmitted once per day via satellite.

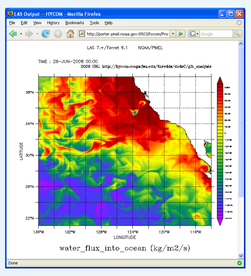
- The CF netCDF conventions provide the ideal format for:
  - Integrating ship and instrument metadata with the underway observations
  - Embedding quality control flags withn the observation file and allowing easy user access to flags and flag definitions (stored as variable attributes)
- We are working towards CF compliance in 2009 to facilitate integration of our data via common server tools (e.g., Live Access Server)

HYbrid Coordinate Ocean Model Consortium (HYCOM)

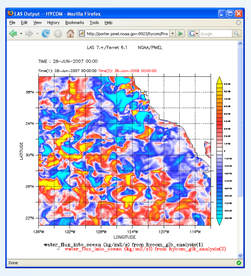
HYCOM is an open source ocean model that uses a hybrid isopycnal-sigma-pressure vertical coordinate to model the vertical structure of the ocean. It is jointly developed by a consortium of national laboratories, universities and private industry including Naval Research Laboratory, University of Miami, Florida state University, National Oceanic and Atmospheric Administration (NOAA) and others. The model source code and various supporting packages are freely available for download from www.hycom.org.

In order to improve usability and inter-operability of HYCOM products software is provided to convert model data to the NetCDF CF-1.0 standards. Standard names from the CF-1.0 name table are used for all variables. These CF compliant NetCDF files can represent both orthogonal curvilinear horizontal grids and rectilinear grids; and in the vertical, both hybrid layers and pressure coordinates. Sample metadata from aHyCOM NetCDF file is shown below.

```
 v {
        String coordinates "Longitude Latitude Date";
        String standard_name
                    "northward_sea_water_velocity";
        String units "m/s";
        Float32 _FillValue 1.2676506E30;
        Float32 valid_range -2.3164985, 3.4698012;
        String long_name " v-veloc. [60.5H]";
    }
    temperature {
        String coordinates "Longitude Latitude Date";
        String standard_name
                  "sea_water_potential_temperature";
        String units "degC";
        Float32 _FillValue 1.2676506E30;
        Float32 valid_range -9.210735, 33.565163;
        String long_name "  temp [60.5H]";
    }
NC_GLOBAL {
        String Conventions "CF-1.0";
        String title "HYCOM GLBa0.08";
        String institution "Naval Research Laboratory";
        String source "HYCOM archive file";
        String experiment "60.5";
        String history "archv2ncdf2d";
```



Live Access Server output of water flux into the ocean off the west cost of the United States, from the HYCOM model

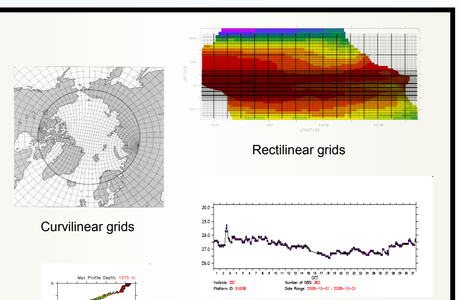


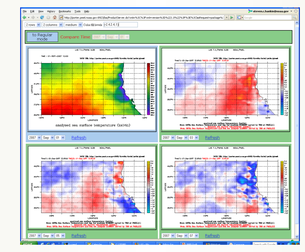
Difference from LAS of water flux into the ocean between June 28, 2007 and June 28, 2008

Goals of the CF Conventions

- Locate data in space-time
- Facilitate processing and graphics
- Identify parameters sufficiently to enable users of data from different sources to decide what is comparable
- Most CF metadata is addressed in a manner not limited to netCDF, and hence can be contained in formats such as XML
- Backwards-compatible with the simpler COARDS standards

CF and grid types

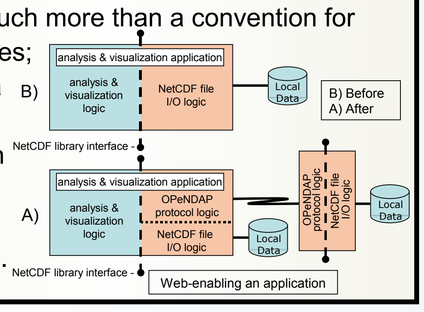




Comparing GHRSST satellite fields with LAS SlideSorter

CF and OPeNDAP

When used in conjunction with the OPeNDAP protocol (http://www.opendap.org/), CF becomes much more than a convention for formatted files; it provides a complete specification for network access to remote data.



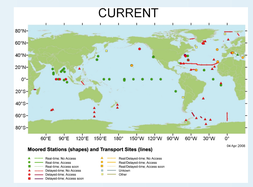
CF Governance

- CF Governance Panel established
- Leadership by two working committees:
  - CF Conventions
  - CF Standard Names
- Committee work done via email and archived web discussion at cfconventions.org
- WCRP/WGCM has been asked to assume responsibility for stewardship
- WCRP/WGNE has been invited to appoint representation on CF Governance Panel

OCEANSites and ARGO

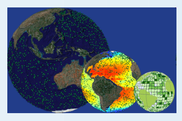
The OceanSITES program is the global network of open-ocean sustained timeseries sites, called ocean reference stations, being implemented by an international partnership of researchers. OceanSITES provides fixed-point timeseries of various physical, biogeochemical, and atmospheric variables at different locations around the globe, from the atmosphere and sea surface to the seafloor. The program's objective is to build and maintain a multidisciplinary global network for a broad range of research and operational applications including climate, carbon, and ecosystem variability and forecasting and ocean state validation.

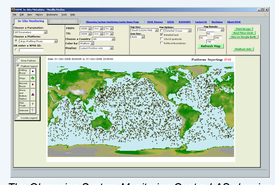
OceanSITES uses the NetCDF (network Common Data Form) system, a set of software libraries and machine-independent data formats. Our implementation of NetCDF is based on the community-supported Climate and Forecast (CF) specification, which supplies a standard vocabulary and some metadata conventions.



OceanSITES has several more restrictions than the CF standard – e.g. variable naming, file naming and limited coordinate systems. These are intended to make it easier to share in-situ data, to make it simpler for the GDACs to aggregate data from multiple sites, and to ensure that the data
can be created and understood by the basic NetCDF utilities.

An OceanSITES data file contains measurements such as temperature and salinity, continuously performed at different levels on a platform (e.g. mooring), as well as meteorological or other parameters recorded at the site, derived variables associated with the site, and complete location, time, and provenance information.





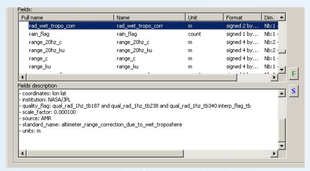
The Observing System Monitoring Center LAS showing Argo platforms reporting during the month of October, 2008

Full resolution Argo profile data, together with quality flags, metadata, and technical data are encoded into netCDFand sent to the Global Data Centres. All data are reported, regardless of the data quality flag assigned by the QC process as well as the data quality flags. In addition, the netCDF file format has a History section used to record which agency carried out which actions on what date and to record what QC tests were performed and which were failed at each Centre. Finally, the netCDF format stores data measured on pressure levels. -- Argo Data Management handbook

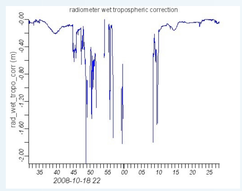
"When Argo format was designed CF wasn't a standard in the In-Situ world. Things have changed and migrating Argo to be CF compliant will be studied at the next Argo format change : considering the large amount of data and data centers involved format changes are not frequent in Argo and we can make a format change only to be CF compliant..." -- Sylvie Pouliquen:

CLS Space Oceanography

Browsing through Jason-2 datasets
The Jason-2 mission, supported by an international partnership of EUMETSAT, the French CNES, and the US agencies NASA and NOAA, collects global ocean surface data. Jason-1 has transitioned from 'classical' binary to netCDF for data dissemination. Moreover, CNES, EUMETSAT, NASA and NOAA decided to utilize CF for Jason-2 datasets. This eases the use of Jason-2 data by end users and increases the potential benefit for network-based applications.



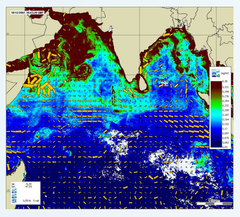
Browsing Jason-2 Interim Geophysical Data Records using BRAT(*) tool. As the Jason-2 dataset includes data from many sources (measured by onboard instruments, or coming from models) it is a good 'exercise in style' of dataset description (here the wet tropospheric correction from the radiometer)



A quicklook on wet tropospheric correction using ncBrowse (**) tool.

THEmatic Maritime Information System (THEMIS)

THEMIS is a map-based application, dedicated to the visualization and management of a wide variety of oceanographic and meteorological gridded datasets, from remote sensing and forecasting systems: satellite altimetry, ocean colour, sea surface or sub-surface temperature, surface currents, atmospheric pressure and winds. THEMIS oceanographic data are stored in netCDF-CF files and provided through an OPeNDAP- THREDDS data server.



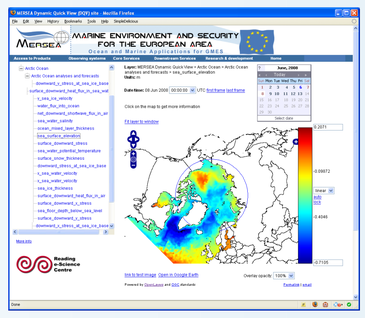
An illustration of the software with plots of chlorophyll concentration with surimposed sea surface geostrophic velocities.

Using CF combined with THREDDS eased the process of automatic acquisition of heterogeneous datasets and visualization. Not only a way to share ocean information through a network, CF is a crucial step towards user friendly ways to discover oceanography.

Who is using CF

- Widely used and accepted in the climate community
  - The foundation for model inter-comparisons in IPCC/AR4 (World Climate Research Programme's Coupled Model Intercomparison Project phase 3 (CMIP3) used by Intergovernmental Panel on Climate Change Working Group 1 for Assessment Report 4)
  - Paleoclimate Modeling Intercomparison Project (PMIP), Hemispheric Transport of Air Pollution (HTAP), regional groups, EU-funded ENSEMBLES prediction system for climate change, . . .
  - Planned for model archives of the next IPCC cycle (CMIP5/AR5)
- Widely adopted in other netCDF archives for atmosphere, oceans, and surface data: ESMF, GFDL, Hadley Centre, NCAR, NOAA, . . .
- Supported by numerous analysis and visualization software packages, file management utilities, and remote data access services via the OPeNDAP protocol.

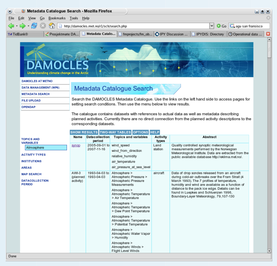
Marine Environment and Security for the European Area (MERSEA)



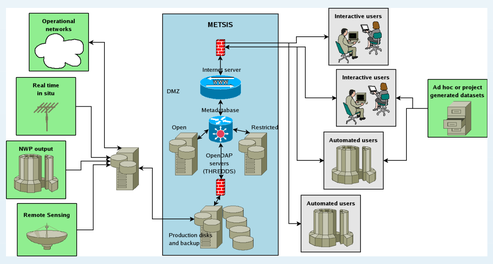
MERSEA Dynamic Quick View website showing a forecast of Arctic Ocean sea surface height. The data's polar stereographic projection is encoded as CF attributes, allowing the data to be correctly geolocated

The CF standard brings huge benefits in the area of code re-use. Scientists and tool developers usually do not have to write complex code to perform complex and error-prone tasks such as the interpretation of numerical grids. Such code is provided by reusable libraries and applications such as the Java NetCDF libraries, Ferret and CDAT. As an example, the MERSEA Dynamic Quick View data visualization website (http://www.resc.reading.ac.uk/mers) interactively displays oceanographic analysis and forecast data from CF-compliant NetCDF files and OPeNDAP servers. Adoption of the CF standard by data providers and software developers has greatly reduced the cost of developing this system, and also means that the process of adding new datasets to this website is straightforward.

DAMOCLES
Developing Arctic Modeling and Observing Capabilities for Long-term Environmental Studies



DAMOCLES is an integrated ice-atmosphere-ocean monitoring and forecasting system for observing, understanding and quantifying climate changes in the Arctic. DAMOCLES is specifically concerned with the potential for a significantly reduced sea ice cover, and the impacts this might have on the environment and on human activities. The data collected and produced by DAMOCLES cover the range from instrumental observations (e.g. CTD, ITP, Gliders) through numerical modeling output to human sciences.

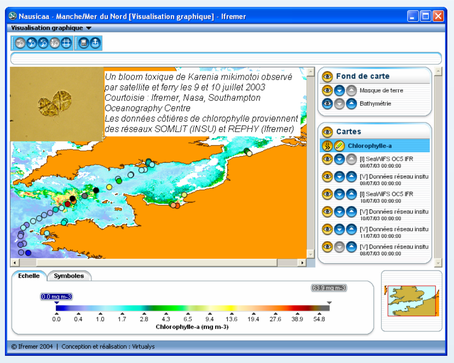


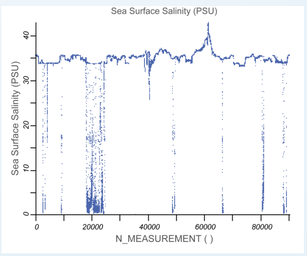
The data management system of DAMOCLES is based upon NetCDF-CF with some DAMOCLES specific extensions. The MERSEA and CORIOLIS CF file formats were used as the starting point. The data management system provides online access both to metadata and data. Metadata are served through a PostGRES database(https://wiki.met.no/metamod/start) and data through an OPeNDAP-THREDDS Data Server. Files may be uploaded using NetCDF or CDL. Uploaded files are checked for compliance and the user receives a note if compliance is not found. Most users are able to generate NetCDF-CF without direct assistance. The DAMOCLES data management system is available at http://damocles.met.no/data_management/, more information on the DAMOCLES project is available at http://www.damocles-eu.org/.

IFREMER
Insitute Français de Recherch pour l'exploitation de la Mer

The Global Ocean Surface Underway Data (GOSUD)* Project is an IOC program designed to collect, process, store and distribute underway surface ocean measurements (mostly salinity and temperature) from ships of opportunity and research vessels. Coriolis hosts the GDAC for GOSUD and NOAA/NODC mirrors it. In both cases, netCDF serves as the primary format for data distribution.

One of the action items on data management out the 2nd GOSUD/SAMOS Workshop** (Seattle, June 10-12 2008) focuses on ensuring that the contents of the NetCDF files being used for data exchange and processing, comply with the CF Metadata Convention. This enhancement will help to improve and assess the metadata consistency between the GOSUD netCDF format and a future BUFR template for TRACKOB data.





Salinity measurements taken by the oceanographic ship Marion Dufresne during 2008. The ncBrowse application has been used to read and display the netCDF file.

On one hand, CF convention helps users to homogeneously read these datasets with their usual scientific tools (IDL, matlab, python...) and on the other hand, it helps systems to process loads of data in a standardized way so that they can provide advanced services to users (online subsetting, viewing, data mining...).

In the future, a forum, an intranet wiki web site and checking tools will provide to IFREMER's different data providers a dedicated support for the netCDF/CF features application (climatological times, geographical projection and boundaries...).

At IFREMER the CF convention is a great support for netCDF file format definition for datasets coming from numerical models (PREVIMER - http://www.previmer.org/), satellite (CERSAT - http://cersat.ifremer.fr/) and in-situ (CORIOLIS : http://www.coriolis.eu.org/) observations. In physical oceanography the CF standard names for parameters are widely used and serve as a federating thesaurus.